

# Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*

Christina Kahramanoglou<sup>1</sup>, Aswin S. N. Seshasayee<sup>2</sup>, Ana I. Prieto<sup>1</sup>, David Ibberson<sup>3</sup>, Sabine Schmidt<sup>3</sup>, Jurgen Zimmermann<sup>3</sup>, Vladimir Benes<sup>3</sup>, Gillian M. Fraser<sup>1,\*</sup> and Nicholas M. Luscombe<sup>2,4</sup>

<sup>1</sup>Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP,

<sup>2</sup>EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK, <sup>3</sup>Genomics Core Facility and <sup>4</sup>Genome Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

Received August 21, 2010; Revised September 27, 2010; Accepted September 28, 2010

## ABSTRACT

Nucleoid-associated proteins (NAPs) are global regulators of gene expression in *Escherichia coli*, which affect DNA conformation by bending, wrapping and bridging the DNA. Two of these—H-NS and Fis—bind to specific DNA sequences and structures. Because of their importance to global gene expression, the binding of these NAPs to the DNA was previously investigated on a genome-wide scale using ChIP-chip. However, variation in their binding profiles across the growth phase and the genome-scale nature of their impact on gene expression remain poorly understood. Here, we present a genome-scale investigation of H-NS and Fis binding to the *E. coli* chromosome using chromatin immunoprecipitation combined with high-throughput sequencing (ChIP-seq). By performing our experiments under multiple time-points during growth in rich media, we show that the binding regions of the two proteins are mutually exclusive under our experimental conditions. H-NS binds to significantly longer tracts of DNA than Fis, consistent with the linear spread of H-NS binding from high- to surrounding lower-affinity sites; the length of binding regions is associated with the degree of transcriptional repression imposed by H-NS. For Fis, a majority of binding events do not lead to differential expression of the proximal gene; however, it has

a significant indirect effect on gene expression partly through its effects on the expression of other transcription factors. We propose that direct transcriptional regulation by Fis is associated with the interaction of tandem arrays of Fis molecules to the DNA and possible DNA bending, particularly at operon-upstream regions. Our study serves as a proof-of-principle for the use of ChIP-seq for global DNA-binding proteins in bacteria, which should become significantly more economical and feasible with the development of multiplexing techniques.

## INTRODUCTION

Transcription in bacteria is controlled by a combination of DNA sequence, topology and a range of *trans*-acting factors (1). The best-studied *trans*-acting regulators are transcription factors (TFs) which modulate transcription by promoting or inhibiting the interaction of the DNA-dependent RNA-polymerase with promoter regions. Bacterial regulators are broadly classified into global and local (2) based primarily on the number of genes that they target for regulation: notable among the former are a subset of ‘nucleoid-associated proteins’ (NAPs) (3).

Many NAPs alter the topology of the bound DNA by bending, wrapping or bridging it (3,4). This has multiple effects on the bacterial cell, among which is transcriptional regulation. Analysis of 12 types of NAPs present in

\*To whom correspondence should be addressed. Email: gmf25@mole.bio.cam.ac.uk  
Correspondence may also be addressed to Nicholas M. Luscombe. Email: luscombe@ebi.ac.uk

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

*Escherichia coli* showed that they differ in their expression during the growth phase (5), the degree of sequence specificity (6) and the capacity for post-translational modifications. Two of the best-studied NAPs, H-NS and Fis, display peak expression during exponential growth and bind to specific DNA sequences and structures. Neither protein is known to be post-translationally modified.

H-NS is a global repressor of transcription in enterobacteria. It acts as a 'genome sentinel' (7) by suppressing the transcription of horizontally-acquired genes, thus providing a fitness benefit to *Salmonella* grown under laboratory conditions (8). It is expressed throughout the growth phase, but shows maximal expression during exponential growth (5), though conflicting data—that H-NS expression is constant across the growth phase—have been presented for *Salmonella* (9). H-NS displays sequence-specific binding and simultaneously affects chromosome structure and transcription by forming DNA–H-NS–DNA bridges, so reinforcing platonically supercoiled structures (10,11).

Fis is a versatile protein which affects multiple processes including transcription, replication and recombination (12). In contrast to H-NS, the expression of Fis peaks during exponential phase, but decreases to undetectable levels in stationary phase (5); therefore, it is thought to be an important player in controlling the growth transition to stationary phase (13). Currently available information in RegulonDB database (14) indicates that, as a TF Fis can activate as well as repress gene expression. On binding, Fis introduces an interwound and branched structure in the DNA where a branch is defined as 'a separate DNA lobe containing at least one intrinsic cross-over'; these structures may be associated with regions of high transcriptional activity (15). Fis influences the distribution of DNA topoisomers in a population of cells: for example *fis* deletion leads to a decreased proportion of cells with low negative supercoiling in stationary phase (13), which might have an impact on stationary phase gene expression.

Analysis of general trends for transcriptional control by Fis and H-NS have generally been performed using compilations of data from small-scale experiments (16). Recently, the use of chromatin immunoprecipitation (ChIP), followed by DNA hybridization to genome-tiling microarrays has led to a systematic and relatively less biased identification of genomic loci physically associated with these proteins, primarily in mid-exponential phase of growth. The only study to have investigated both proteins simultaneously—using microarrays with probes tiled at 160bp resolution—showed that there is significant overlap between the genes targeted by the two proteins (17). There are two contrasting models of how H-NS represses transcription: Lucchini and colleagues proposed that H-NS inhibits the initial RNA-polymerase-DNA interaction (8), whereas Grainger and co-workers and Oshima et al. demonstrated that the polymerase is trapped at the promoter (17,18). For Fis, the majority of bound genes were shown not to change in expression in a *fis* deletion strain (19), which is intriguing given that Fis is considered to be a global regulator of transcription.

Despite the above studies, we do not know whether and how the binding of these proteins to the DNA varies across the growth phase. This is particularly important since their expression levels are known to change substantially during growth. It has been previously suggested that H-NS might act both as a canonical TF and as a silencer of gene expression (20); however, the distinction between these two modes of H-NS function have not been described on a genomic scale. Finally, given prior observations of limited overlap between genes bound by Fis and those that change in expression in a deletion strain (19), the genome-scale impact of Fis–DNA interactions on gene expression remains poorly understood.

Here, we present an investigation of genomic loci bound by Fis and H-NS in *E. coli* K12 using ChIP followed by high-throughput sequencing, instead of microarray hybridization, of the immunoprecipitated DNA (ChIP-Seq) (21). Improvements in sequencing have revolutionized genomics by providing a platform for quantifying nucleic acid concentrations that affords higher dynamic range, higher resolution and lower false positive rates (22,23). These are now being used extensively to investigate protein–nucleic acid interactions in eukaryotes (21,24–27). In bacteria, however, their use has been largely limited to whole-genome sequencing and transcriptomic analysis (28–33), though transcriptome-level investigations have been extended using immunoprecipitation-based interrogation of protein–RNA interactions in *Salmonella* (34). Recently, a ChIP-Seq-based analysis of a *Mycobacterium tuberculosis* TF DosR, which binds to ~25 loci on the genome, was published (35). To our knowledge, ours is the first detailed genome-scale interrogation of protein–DNA interactions, for any global DNA binding protein in bacteria, using high-throughput sequencing. In addition to providing a proof-of-principle for the use of this new technology for bacteria, we perform our study at multiple time-points during growth in rich medium, thus generating new insights into how these proteins function under different cellular conditions. Further, by analysing our data in conjunction with gene expression and RNA-polymerase–DNA interaction data we provide new interpretation of the regulatory functions of these proteins.

## MATERIALS AND METHODS

### Strains and general growths conditions

The *E. coli* K-12 MG1655 bacterial strains used in this work are the following: *E. coli* MG1655 (F-lambda-ilvG- rfb-50 rph-1); MG1655 *Δhns* (*Δhns*::Kan<sup>r</sup>); MG1655 *Δfis* (*Δfis*::Kan<sup>r</sup>); MG1655 *hns*-FLAG (*hns*::3xFLAG::Kan<sup>r</sup>); MG1655 *fis*-FLAG (*fis*::3xFLAG::Kan<sup>r</sup>). Luria-Bertani (0.5% NaCl) broth and agar (15 g/liter) were used for routine growth. Where needed, ampicillin, kanamycin and chloramphenicol were used at final concentrations of 100, 30 and 30 µg/ml, respectively.

### Construction of *E. coli* MG1655 knock-outs and FLAG-tagged strains

Disruption of *hns* and *fis* genes in the *E. coli* chromosome was achieved by the  $\lambda$  Red recombination system, previously described by Datsenko and Wanner (36). Primers designed for this purpose are shown in Supplementary Data 19. Sets of additional external primers were used to verify the correct integration of the PCR fragment by homologous recombination (Supplementary Data 19).

The 3xFLAG epitope was added at the C terminus of the H-NS and Fis protein by a PCR-based method with plasmid pSUB11 as template (37). Primers used for introducing the 3xFLAG tag are shown in Supplementary Data 19. The tagged construct was then introduced onto the chromosome of *E. coli* MG1655 using the  $\lambda$  Red recombinase system (36). At each stage, DNA and strain constructions were confirmed by PCR and/or sequencing. This approach resulted in the introduction of a kanamycin resistance cassette in the chromosome downstream of the tagged gene. The cassette can be removed by FLP-mediated site-specific recombination (36), although this was not done for the experiments described here. In all cases, the complete functionality of the 3xFLAG-tagged version of the proteins was tested.

### RNA extraction and microarrays procedures

To prepare cells for RNA extraction, 100 ml of fresh LB was inoculated 1:200 from an overnight culture in a 250 ml flask and incubated with shaking at 180 rpm in a New Brunswick C76 waterbath at 37°C. Two biological replicates were performed for each strain and samples were taken at early-exponential, mid-exponential, transition-to-stationary and stationary phase. The cells were pelleted by centrifugation (10000g, 10 min, 4°C), washed in 1× PBS and pellets were snap-frozen and stored at -80°C until required. RNA was extracted using Trizol Reagent (Invitrogen) according to the manufacturer's protocol until the chloroform extraction step. The aqueous phase was then loaded onto mirVana™ miRNA Isolation kit (Ambion Inc.) columns and washed according to the manufacturer's protocol. Total RNA was eluted in 50 µl of RNAase free water. The concentration was then determined using a Nanodrop ND-1000 machine (NanoDrop Technologies), and RNA quality was tested by visualization on agarose gels and by Agilent 2100 Bioanalyser (Agilent Technologies).

For the generation of fluorescence-labelled cDNA the FairPlay III Microarray Labelling Kit (Stratagene) was used. Briefly, 1 µg of total RNA was annealed to random primers, and cDNA was synthesized in a reverse transcription reaction with an amino allyl modified dUTP in the presence of 1 µg of Actinomycin D. The amino allyl labelled cDNA was then coupled to a Cy3 dye (GE Healthcare) containing a NHS-ester leaving group. The labelled cDNA was hybridized to the probe DNA on custom Agilent microarrays by incubating at 65°C for 16 h. The unhybridized labelled cDNA was removed and the hybridized labelled cDNA was visualized using an Agilent Microarray Scanner. Note that we performed a one-colour experiment on the Agilent array.

### RT-PCR for validation

To validate the results of the microarray analysis, quantitative reverse-transcriptase PCR (qRT-PCR) was carried out using specific primers to the mRNA targets showing up- or down-regulation, and control targets not showing differential expression. RNA was extracted as described above from wild-type,  $\Delta hns$  and  $\Delta fis$  cells and 30 ng total RNA was used with the Express One-Step SYBR GreenER kit (Invitrogen) according to the manufacturer's guidelines, using a MJ Mini thermal cycler (Bio-Rad).

### ChIP

ChIP was performed as previously described (38) with some modifications to the protocol. Cells were grown aerobically at 37°C to the desired OD600 and formaldehyde was added to a final concentration of 1%. After 20 min of incubation, glycine was added to a final concentration of 0.5 M to quench the reaction and incubated for a further 5 min. Cross-linked cells were harvested by centrifugation and washed twice with ice-cold TBS (pH 7.5). Cells were resuspended in 1 ml of lysis buffer [10 mM Tris (pH 8.0), 20% sucrose, 50 mM NaCl, 10 mM EDTA, 20 mg/ml lysozyme and 0.1 mg/ml RNase A] and incubated at 37°C for 30 min. Following lysis, 3 ml immunoprecipitation (IP) buffer [50 mM HEPES-KOH (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% sodium dodecyl sulphate (SDS) and PMSF (final concentration 1 mM)] was added and the lysate passed through a French pressure cell twice. Two microlitres of aliquots were removed and the DNA sheared to an average size of ~200 bp using a Bioruptor (Diagenode) with 30 cycles of 30 s on/off at high setting. Insoluble cellular matter was removed by centrifugation for 10 min at 4°C, and the supernatant was split into two 800 µl aliquots. The remaining 400 µl was kept to check the size of the DNA fragments.

Each 800 µl aliquot was incubated with 20 µl Protein A/G UltraLink Resin (Pierce) on a rotary shaker for 45 min at room temperature to get rid of complexes binding to the resin non-specifically. The supernatant was then removed and incubated with either no antibody (mock-IP), FLAG mouse monoclonal antibody (Sigma-Aldrich) or RNAP  $\beta$  subunit mouse monoclonal (NeoClone) and 30 µl Protein A/G UltraLink Resin, pre-incubated with 1 mg/ml bovine serum albumin (BSA) in TBS, on a rotary shaker at 4°C overnight (FLAG antibody) or at room temperature for 90 min (RNAP  $\beta$  subunit antibody). Samples were washed once with IP buffer, twice with IP buffer + 500 mM NaCl, once with wash buffer [10 mM Tris (pH 8.0), 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet P-40 and 0.5% sodium deoxycholate] and once with TE (pH 7.5). Immunoprecipitated complexes were eluted in 100 µl elution buffer [10 mM Tris (pH 7.5), 10 mM EDTA and 1% SDS] at 65°C for 20 min.

Immunoprecipitated samples and the sheared DNA following the Bioruptor were de-crosslinked in 0.5× elution buffer containing 0.8 mg/ml Pronase at 42°C for 2 h followed by 65°C for 6 h. DNA was purified using a PCR purification kit (QIAGEN). Prior to sequencing, the



DNA fragment sizes were checked and gene-specific quantitative PCR (qPCR) was carried out.

### Real-time qPCR

To measure the enrichment of the Fis, H-NS or RNAP-binding targets in the immunoprecipitated DNA samples, real-time qPCR was performed using a MJ Mini thermal cycler (Bio-Rad). One microlitre of IP or mock-IP DNA was used with specific primers to the promoter regions (primer sequences are available upon request) and Quantitect SYBR Green (QIAGEN).

### Library construction and Solexa sequencing

Prior and post library construction, the concentration of the immunoprecipitated DNA samples was measured using the Qubit HS DNA kit (Invitrogen). Library construction and sequencing was done using the ChIP-Seq Sample Prep kit, Reagent Preparation kit and Cluster Station kit (Illumina). Samples were loaded at a concentration of 10 pM.

### Public data sources

The *E. coli* K12 MG1655 genome was downloaded from the KEGG database (39). Annotations of gene coordinates were obtained from Ecocyc 11.5 database (40). Literature-derived transcriptional regulatory network, including known targets for Fis and H-NS, for *E. coli* K12 was obtained from RegulonDB 6.2 database (14). Targets of the global transcriptional regulator CRP were downloaded from RegulonDB 6.2 and augmented, where required, with additional targets identified by Grainger and colleagues (41). Genomic regions with atypical composition of higher-order oligonucleotides—and thus putatively corresponding to horizontally-acquired DNA—were identified using the Alien Hunter software (42). Lists of genes identified as bound by H-NS or Fis in previous high-resolution tiling microarray studies were downloaded from the respective publications (17–19). Protein occupancy domains of *E. coli* were downloaded from Vora *et al.* (43). Orthologs between *Salmonella enterica* Typhimurium LT2 and *E. coli* K12 MG1655 were obtained from the work of Moreno-Hagelsieb and Latimer (44).

### Analysis of protein binding regions from high-throughput sequencing data

*Identification of protein binding regions on the DNA.* Sequences obtained from the Illumina Genome Analyzer were mapped to both strands of the *E. coli* K12 MG1655 genome using BLAT allowing no gaps and up to two mismatches. Each alignment was extended to 200 bp—the approximate average length of DNA fragments—on the 3' end. Only reads which mapped to a single region of the genome were considered for further analysis. For each base position on the genome, the number of reads that mapped to that position was calculated. The distribution of read counts thus obtained had a sharp peak at a low value followed by a heavy tail. Since this characteristic of the distribution

is similar to that obtained for high-resolution gene expression tiling arrays, we used a procedure adopted earlier for tiling array analysis (45). Briefly, the background was a normal distribution with the following parameters:  $\mu$  = mode (as computed using the 'shorth' procedure in R) of the entire distribution and;  $\sigma = 1.483 \times$  median absolute deviation of all values less than the mean of the entire distribution. This gives a better fit of the empirical distribution than a Poisson distribution of the same  $\mu$  (Supplementary Data 1). The cutoff read count was defined as  $Z = \mu + 3\sigma$ . Any consecutive stretch of DNA where each coordinate had a read count greater than or equal to the cutoff was flagged; pairs of adjacent regions so obtained were merged to give a single binding region if they were separated by <200 bp. Then the number of reads mapped to each binding region, normalized by the total number of reads obtained for that sample, was compared to the corresponding value from the mock-IP using a binomial test, as described in the PeakSeq algorithm (46). Any region giving a Bonferroni-corrected  $P \leq 0.01$  was defined as a bonafide protein binding region. We performed mock-IP only in the mid-exponential phase taking into consideration the following: (i) it has been suggested that a single control library can be used across multiple ChIP-Seq experiments given that these were performed in the same organism under similar fragmentation conditions (47); (ii) qPCR data for mock-IP experiments from our laboratory show minimal and inconsistent variation across time-points.

*Comparison with previously published ChIP-chip datasets.* For this purpose, we downloaded lists of genes identified as bound (either upstream or in genebody) by H-NS and Fis from published tiling microarray studies (17–19). These genes were compared with those which overlap with binding regions identified in our study; here the cut-off for defining an overlap was set at 100 bp. Here, we used the union of genes detected as bound by the protein of interest in early- and mid-exponential phases of growth to partly account for possible differences in the environmental/cellular conditions used in the compared studies.

*Identification of and scanning for sequence motifs.* To identify DNA sequence motifs for the binding of H-NS and Fis, we obtained the sequence of 101 bp of DNA including 50 bp on either side of the summit for each binding region. Here the summit for each binding patch was defined as the base coordinate with the highest read count within that region. The sequences so obtained were scanned for motifs using the MEME software (48) with the following parameters: zero or one motif per sequence; motif width ranging from 6 to 24; searching both strands of the sequences; using a background distribution file containing the mono- and di-nucleotide frequencies of the *E. coli* chromosome. Then the complete sequence of each binding region was searched for the presence of these motifs using the MAST programme (48) with a  $P \leq 0.001$  and using the same background mono- and di-nucleotide frequencies as above. Any definition of a motif in this

work refers to those which were identified within the binding regions.

**Definition of operons bound by the protein of interest.** We used the operon definitions available in RegulonDB 6.2 (14) to identify a set of 2567 lead genes, which are the first genes of each operon. An operon was flagged as being bound by the protein of interest if at least 50 bp of the intergenic region upstream of the operon overlapped with a binding region. For long intergenic regions, only the first 400 bp of the sequence immediately upstream of the operon were used.

**Derivation of RNA-polymerase occupancy from high-throughput sequencing data.** Reads obtained from the sequencing of RNA-polymerase-associated DNA were mapped to the genome and read counts obtained per base position as described above. For each gene, the median read count across all base positions corresponding to the gene body was defined as its occupancy. In addition, for each lead gene, the highest read count in the upstream region was calculated and used as a representation of transcription initiation. Data from each sample were normalized to the total number of reads obtained for that sample and then divided by the corresponding value from the mock-IP.

#### Analysis of gene expression from microarray data

Gene expression analyses were performed on custom-designed isothermal Agilent microarrays containing 10 821 60-mer probes covering 4373 genes. In addition to these sense probes, the array contained 4172 anti-sense probes which were excluded from this analysis. These probes were designed using Array Oligo Selector (49).

Microarray data were processed in Bioconductor using standard procedures. Briefly, array data were background corrected using *normexp* (50). Biological replicates were first normalized using variance stabilization and normalization (VSN) (51). All arrays, across genetic backgrounds, from the same time-point were again normalized together using VSN. Differential expression in the deletion strains compared with the wild-type at the same time-point was called at false discovery rate (FDR)-adjusted *P*-value of 0.01 using the LIMMA package (52); this was performed at the level of individual probes. Any gene was called differentially expressed even if one of the probes corresponding to it passed the above threshold. For direct comparison with operons that are bound by the protein of interest, we used only the list of lead genes that were differentially expressed. 'Absolute' expression level for each probe under a given genetic background and growth phase, where required, was defined as the average value across replicates; this shows a significant correlation with RNA-seq data obtained in our lab during exponential phase of growth (Spearman Rank correlation = 0.73; Supplementary Data 19).

#### Statistical analysis

The Fisher's exact test was used for categorical data. Wilcoxon-Mann-Whitney tests were performed when

comparing distributions. Since the size of the distributions were typically large, we used the *t*-test as well to ensure that the result of our comparisons were significant in both tests. In this article, we report *P*-values from the Wilcoxon test. Unless otherwise stated, a *P*-value cutoff of 0.01 was used to signal statistical significance. Correlation coefficients of read count 'Z' (see above) between two samples were computed at the base resolution, ignoring 'background' coordinates where the Z for both samples were <2. All these tests were carried out in R (www.r-project.org).

#### Accession numbers

All microarray and ChIP-seq data have been submitted to ArrayExpress, and have been assigned the following accession numbers: ChIP-seq: E-MTAB-332; Microarray design: A-MEXP-1866; Microarray raw and normalized data: E-MEXP-2838; RNA-seq data: E-MTAB-387.

## RESULTS

### Growth-phase-specific binding of H-NS and Fis to the *E. coli* chromosome

**Immunoprecipitation-sequencing of genomic DNA cross-linked to H-NS and Fis.** We investigated H-NS- and Fis-binding to the *E. coli* K12 chromosome during early-exponential, mid-exponential, transition-to-stationary and stationary phases of growth in LB medium by ChIP combined with high-throughput sequencing (21,23). As controls, we performed mock-IP experiments in mid-exponential phase in the absence of antibodies to identify non-specifically precipitated DNA.

For each sample, we obtained ~6–15-million reads of 36-nt length, amounting to 50–120-fold coverage of the *E. coli* genome (Table 1). We mapped these reads to both strands of the *E. coli* K12 MG1655 genome sequence and extended the mapping in the 3'-end to 200 bp, which is the approximate average length of the DNA fragments obtained from the immunoprecipitation experiments. To identify bound loci, we calculated the number of reads that mapped to each base-pair in the genome (Figure 1). We expected to see a near-complete representation of the genome in our sequences, irrespective of where the proteins bind; therefore we derived an internal background distribution for each sample as described earlier for tiling microarray data (see 'Materials and methods' section, Supplementary Data 1) (45). The cutoff value for calling binding regions was fixed at three standard deviations above the mean of the background normal distribution (not more than 1% of values within the normal distribution are higher than this cutoff). Any stretch of DNA where each position mapped to more reads than the above-defined cutoff was called a binding region. Then, all binding regions separated by <200 bp were merged; this was performed to counter possible under-sequencing of chromosomal regions of length equal to a single read (22). Finally, binding regions whose read counts did not differ significantly from the mock-IP sample were removed. Selected binding regions

Table 1. Analysis of Fis and H-NS ChIP-Seq raw data, showing number of reads, binding regions and genomic features

	Fis <sup>a</sup>			H-NS		
	Early-exponential	Mid-exponential	Stationary	Early-exponential	Mid-exponential	Stationary
Number of reads	6170 274	7285 416	14 576 723	10017 793	6574 797	11 908 288
Number of mappable reads	4614 638 (74.8%)	6123 662 (84.1%)	14 133 453 (97.0%)	9 319 380 (93.0%)	6324 071 (96.2%)	11 308 283 (95.0%)
Number of unique reads <sup>b</sup>	4461 734 (96.7%)	5945 924 (97.1%)	13 775 767 (97.5%)	9 060 279 (97.2%)	6154 620 (97.3%)	11 004 456 (97.3%)
Number of reads per base						
Mean	192.8	256.9	595.3	391.5	266.0	475.5
Mode <sup>c</sup>	166.3	212.4	233.2	222.9	124.4	272.0
Number of bases within binding regions	444 427 (9.6%)	520 151 (11.2%)	944 735 (20.4%)	648 991 (14.0%)	772 282 (16.6%)	809 040 (17.4%)
Number of binding regions	1246	1464	537	444	458	547
Average length of binding regions (base pairs)	356.7	355.3	1759.3	1461.7	1686.2	1479.1
Number of operons targeted <sup>d</sup>	516	649	769	498	597	634

<sup>a</sup>Fis is not expressed at detectable levels in transition-to-stationary and stationary phases of growth. Attempts at immunoprecipitating DNA cross-linked to Fis during these phases of growth failed to return enough DNA to perform a sequencing reaction.  
<sup>b</sup>This is the number of reads that were mapped to a single region of the *E. coli* K12 MG1655 genome. These reads were used for all further analyses. The percentage is out of all mappable reads.  
<sup>c</sup>This was used as the mean of the normal distribution representing the background.  
<sup>d</sup>Any operon was considered bound by the protein concerned if at least 50 bp of any binding region overlapped with the upstream intergenic region of the operon (subject to a maximum of 400-bp upstream).

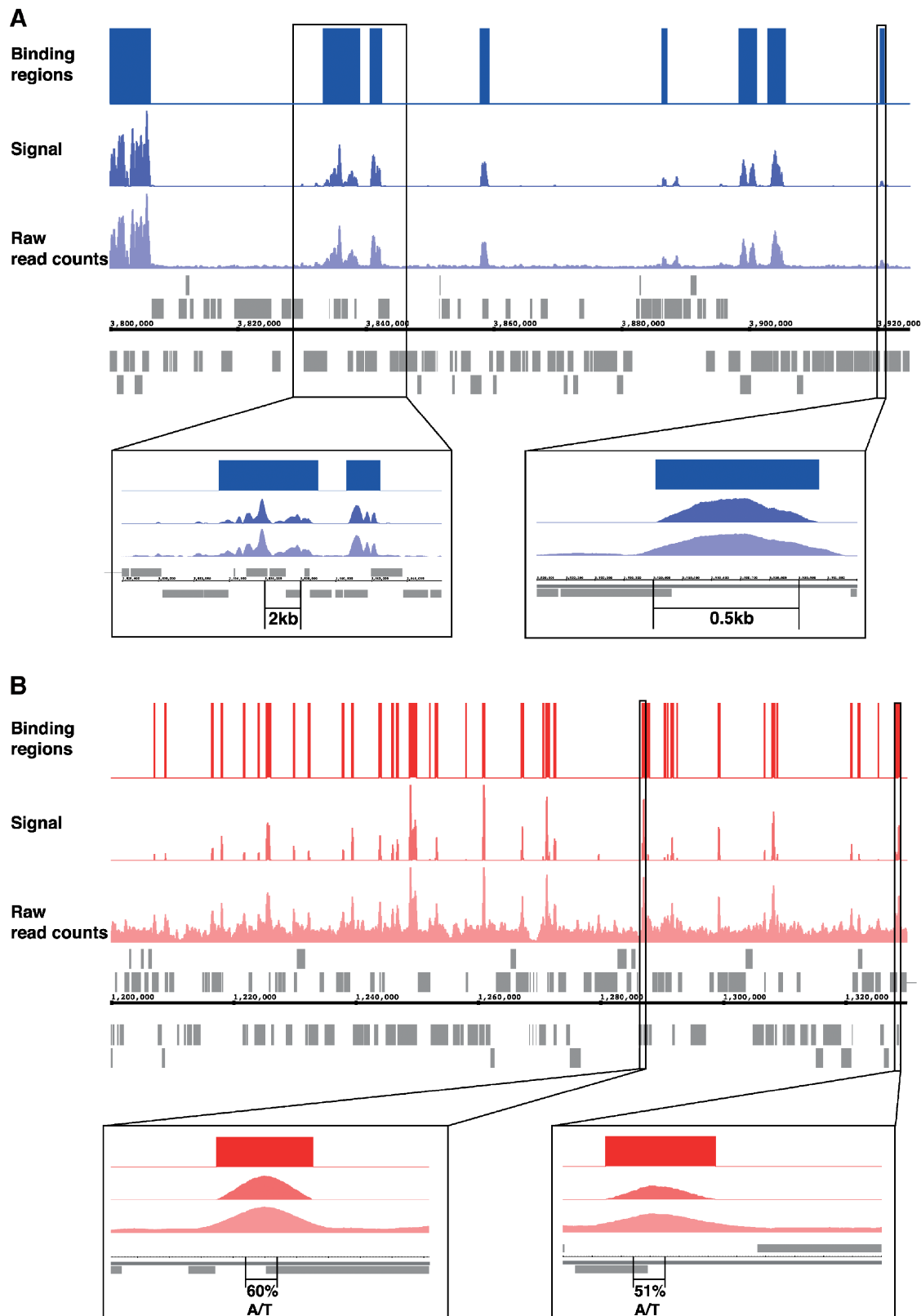
were verified using quantitative PCR (Supplementary Data 19).

*Comparison with previously published high-throughput datasets.* First, we compared our dataset with previously published ChIP-chip data. Here we note that cross-study comparisons are not straightforward owing to differences in experimental conditions and platforms, analysis procedure and the manner in which data are presented. However where possible, we use published lists of bound genes and raw binding signals as points of comparison.

We compared our data for H-NS (combining early- and mid-exponential phase data) with that from a tiling microarray-based study by Oshima and co-workers (18). A large majority of genes (75%; Fisher's exact test  $P < 10^{-50}$ ) flagged as bound by H-NS in the above study overlap with binding regions (by at least 100 bp) we identify; just over a quarter of genes (27%) detected in our study are not identified by Oshima and colleagues.

We then compared ChIP-chip data for Fis by Cho and colleagues (19) with our data. Overall, binding regions that we identify display significantly higher fluorescence intensities in the above dataset than randomly picked regions (Supplementary Data 2). This is despite the fact that Cho et al. performed their experiments in M9 plus glucose whereas ours was carried out in LB without supplemented sugars. Over two-thirds (67%; Fisher's exact test  $P < 10^{-50}$ ) of genes bound by Fis in the Cho dataset overlap with binding regions identified here. However, we detect a significantly larger number of bound genes, with binding either in the gene body or in upstream regions (1592 genes, compared with 894 genes in the Cho dataset). Even at a more stringent threshold for identifying binding regions, we identify more bound genes (1006 genes) than Cho *et al.* with a recovery of 53% (Fisher's exact test  $P < 10^{-50}$ ).

We then compared our dataset with the lower resolution (160-bp resolution) ChIP-chip study of Grainger and colleagues (17). For H-NS, there is excellent agreement in binding signals between the two datasets. However the overlap at the gene level is poor (39% of genes flagged as bound by H-NS by Grainger and colleagues are recovered here; Supplementary Data 2). We believe that the poor overlap at the gene level is a consequence of assumptions made in assigning binding regions to target genes; this could have been exacerbated by the lower resolution of the Grainger study (David Grainger, personal communication). Remarkably for Fis, there is no similarity between the two datasets either at the level of binding signals or bound genes (31% of bound genes in the Grainger dataset are recovered). This might be a consequence of differences in experimental conditions, which might affect Fis more than H-NS because of the former's link with catabolite repression (53,54); in fact, we observe a statistically significant overlap (Fisher's exact test  $P = 9.8 \times 10^{-6}$ ) between operons bound in their upstream regions by Fis (but not H-NS) in our study and publicly-available targets of CRP (14,41), the global regulator of catabolite repression. However, we note that there is only a limited correlation in binding signals between the studies of Cho *et al.* and Grainger



**Figure 1.** ChIP-Seq of H-NS and Fis in *E. coli* K12: tracks showing binding signal (raw number of reads mapping to each basepair); signal in the form of  $z$ -scores compared with the background normal distribution; binary representation for coordinates representing binding regions) for H-NS (blue, A) and Fis (red, B). In the track showing the 'signal', all  $z$ -scores were subtracted by three to remove background; negative signals are not shown. The insets for H-NS show 'long' and 'short' binding regions; those for Fis show binding regions with 'high' and 'low' A/T contents.



*et al.* despite the fact that both studies were performed in minimal media with a difference only in the identity of the carbon source used (glucose and fructose, respectively; Supplementary Data 2). It is possible that Fis binds to the *E. coli* genome extensively and that each study sampled only a subset of binding sites: this might be substantiated by the fact that the background signal is higher for Fis than H-NS (Figure 1; Stephen Busby, personal communication).

Finally, we compared the lists of genes identified as H-NS targets in *S. enterica* Typhimurium (8) with our data. A majority (58%) of genes bound by H-NS in *Salmonella* do not have orthologs in *E. coli* (Fisher's exact test,  $P = 5.2 \times 10^{-36}$ ). These genes are probably horizontally acquired, and are exemplified by the H-NS-mediated regulation of the *Salmonella*-specific pathogenicity islands such as Spi-1 and Spi-2 which have been horizontally-acquired. Similarly, ~46% of genes with H-NS binding in *E. coli* do not have orthologs in *Salmonella*. Therefore, the targets of H-NS are substantially different between *E. coli* and *Salmonella*. We note here that over 75% of the conserved H-NS targets in *Salmonella* are bound by H-NS in *E. coli*; this proportion is similar to the agreement between two independent studies of H-NS targets in *E. coli* (see above).

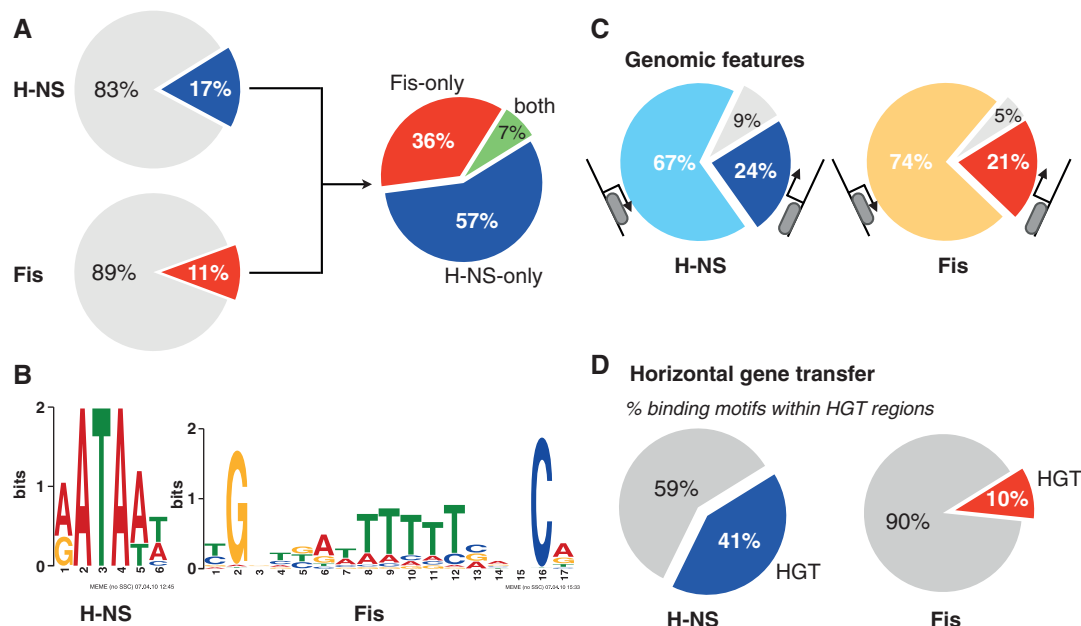
**DNA-binding profiles of H-NS and Fis in mid-exponential phase.** We focus on the DNA-binding profiles of H-NS and Fis in mid-exponential phase (Figure 1 and Table 1). H-NS binds to ~17% of the genome in terms of basepairs, whereas Fis binds to ~11%, distributed over 458 and 1464 discrete binding regions, respectively (Figure 2A).

In contrast to observations of Grainger and colleagues (17)—made under substantially different growth conditions—we find little overlap between Fis and H-NS-binding regions (Figure 2A). In fact, across the genome, there is a significant negative correlation between the binding signals for H-NS and Fis.

H-NS binds to longer tracts of DNA than Fis (averages of 1686 and 355 bp for H-NS and Fis, respectively; Wilcoxon test,  $P < 10^{-50}$ ; Figure 3A and B; Supplementary Data 3). The observed length distribution for H-NS is in line with the results of a recent study in *Salmonella* (55). This is consistent with the ability of H-NS to form long oligomers, extending from high affinity nucleation sites to flanking lower affinity sites (10,56).

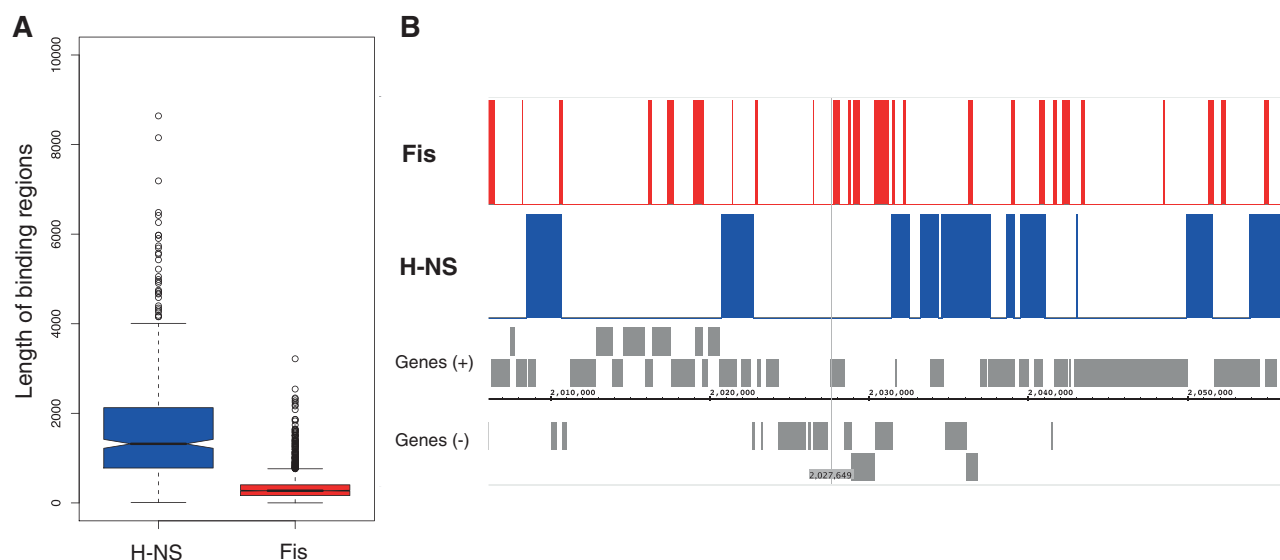
The H-NS binding motif (57), defined by enriched oligonucleotide sequences within bound regions, is 5–6 nt in length and comprises only A/T nucleotides (Figure 2B; Supplementary Data 4). This motif is present in 96% of all binding regions at an average of 19.9 occurrences per region. In agreement with published results (19), the ~15-nt Fis motif consists of an A/T tract flanked by highly conserved G/C on either side (Figure 2B). This motif is present in 91% of binding regions at an average of 2.3 occurrences per region. Note that we differentiate between binding *regions* and *motifs*: whereas binding regions are empirically identified by our experiments, motifs represent the computationally identified sequences that fall within our binding regions.

On average, 18 and 17% of all basepairs covered by H-NS and Fis binding regions—corresponding to 24 and 21%, respectively, of binding motifs—fall within intergenic



**Figure 2.** Binding of H-NS and Fis to the *E. coli* K12 chromosome. (A) Proportion of the genome that falls within the binding regions of H-NS (blue) and Fis (red); grey sectors of the pie charts represent the proportion of unbound regions. (B) Proportion of base pairs that are bound by H-NS only (blue), Fis only (red), or by both (green). (C) Logos representing the most significant motif associated with H-NS and Fis binding regions. (D) Proportion of binding motifs that fall within the following genomic features: operon-upstream regions (blue for H-NS, red for Fis), operon body (light blue, orange), other regions (grey). (E) Proportions of binding motifs that fall within predicted horizontally-acquired DNA (blue for H-NS, red for Fis).





**Figure 3.** Length of H-NS- and Fis-binding regions. (A) Boxplot showing the distributions of the lengths of binding regions for H-NS (blue) and Fis (red) in mid-exponential phase. (B) Genomic browser (Integrated Genome Browser) view of an ~50-kb region of the *E. coli* genome showing H-NS (blue) and Fis (red) binding regions in mid-exponential phase; the grey bars in the bottom panel represent annotated genes on the + (top) and the - (bottom) strands.

regions upstream of predicted operons (Figure 2C; Supplementary Data 5). Given that ~8% of the *E. coli* genome comprises operon-upstream intergenic regions, Fis and H-NS display a preference for binding upstream of operons. Most of the other motifs fall within the body of operons (67 and 74% for H-NS and Fis, respectively). In agreement with previous reports (8,18,58), there is significant enrichment of H-NS (but not Fis) binding across horizontally-acquired regions (Figure 2D).

Finally, we define 597 (23%) and 649 (25%) operons as bound in a regulatory capacity by H-NS and Fis, respectively, based on binding in upstream regulatory sequences (applying a limit of 400 bp). The rest of our discussion is based on the above operons only and not those with protein binding only in the gene body (as included in our comparison with previous studies). Operons targeted by H-NS are enriched for gene functions associated with fimbrial biogenesis (Fisher's exact test,  $P = 5.1 \times 10^{-3}$ ), which expands previous work linking H-NS to the regulation of biofilm formation and motility (59). As expected from prior molecular studies (60), operons bound by Fis show an enrichment for genes involved in translation (Fisher's exact test,  $P = 1.6 \times 10^{-3}$ ). In agreement with the significant overlap of Fis bound genes with CRP targets, carbohydrate metabolism and transport also shows a significant enrichment among Fis targets (Fisher's exact test,  $P = 4.7 \times 10^{-3}$ ).

**Length of H-NS binding regions and their characteristics.** We had noted earlier that genomic regions bound by H-NS tend to be longer than those bound by Fis (Figure 3). In order to investigate systematically the association between the length of H-NS binding regions and genomic features recognized, we classified H-NS binding regions into those that are longer than 1000bp ('L<sub>H-NS</sub>' regions;  $n = 300$  in mid-exponential

phase) and those that are shorter ('S<sub>H-NS</sub>' regions;  $n = 158$ ). We observe that a significantly higher proportion of motifs within S<sub>H-NS</sub> (37%) than L<sub>H-NS</sub> (22%) regions fall in operon-upstream regions (Fisher's exact test,  $P = 1.2 \times 10^{-21}$ ; Supplementary Data 5). This might be expected given the differences in their lengths and the fact that operon-upstream regions have high A/T content (61). Unexpectedly however, the proportion of operon-upstream S<sub>H-NS</sub> motifs is significantly higher than that for Fis motifs as well (Fisher's exact test,  $P = 1.3 \times 10^{-20}$ ).

We also observe that horizontally-acquired genes are significantly enriched in the L<sub>H-NS</sub> group (Supplementary Data 6); this is in accordance with the fact that predicted horizontally-acquired genes are located in long regions of typically higher A/T content than the genomic average (Supplementary Data 7).

Therefore, short H-NS binding regions tend to behave in a manner typical of canonical TFs, where the protein binds upstream of the gene whose expression it regulates. On the other hand, longer binding regions wrap large segments of the chromosome, encompassing both genes and intergenic regions.

**Variable structures of Fis-DNA complexes.** It was previously demonstrated that Fis-DNA complexes adopt variable structures depending on the A/T content of the DNA surrounding the core binding motif (62). The variability in these complexes is manifested by the degree to which the bound DNA is bent, with greater bending in regions of higher A/T content. To investigate this in our data, we defined binding regions in the top quarter of the distribution of A/T contents (101 bp around the summit) as likely to be bent by Fis. The association between the A/T content of the binding region and gene expression is described later.

As intergenic regions tend to have higher A/T contents and intrinsic bending, a greater proportion of motifs falling within high-A/T binding regions are in operon-upstream regions (40 versus 11%; Wilcoxon test,  $P < 10^{-50}$ ; Supplementary Data 5). Furthermore, high A/T Fis-binding regions show significantly greater binding signal than other regions (Wilcoxon test,  $P = 3.5 \times 10^{-9}$ ; Supplementary Data 8); this might reflect the fact that Fis-DNA complexes involving DNA bending dissociate slower than others (62).

*Growth-phase dependent variation in the DNA-binding profiles of H-NS and Fis.* Next, we investigated the variation in H-NS and Fis-binding profiles from the early-exponential to the stationary phases of growth (Figure 4).

For H-NS, we detected similar binding at all four stages of growth (Table 1 and Figure 4). Though previous experiments showed a peak for H-NS protein expression during exponential growth followed by an  $\sim 2\text{--}2.5 \times$  decrease during later stages (5), our western blot experiment showed constant H-NS levels across our experimental conditions (Supplementary Data 9), in agreement with previous results in *Salmonella* (9). For Fis, we identified comparable numbers of binding regions in both early- and mid-exponential phases (Table 1 and Figure 4). In agreement with earlier studies, our western blots (Supplementary Data 9) show that Fis is expressed below detectable levels after exponential growth (5).

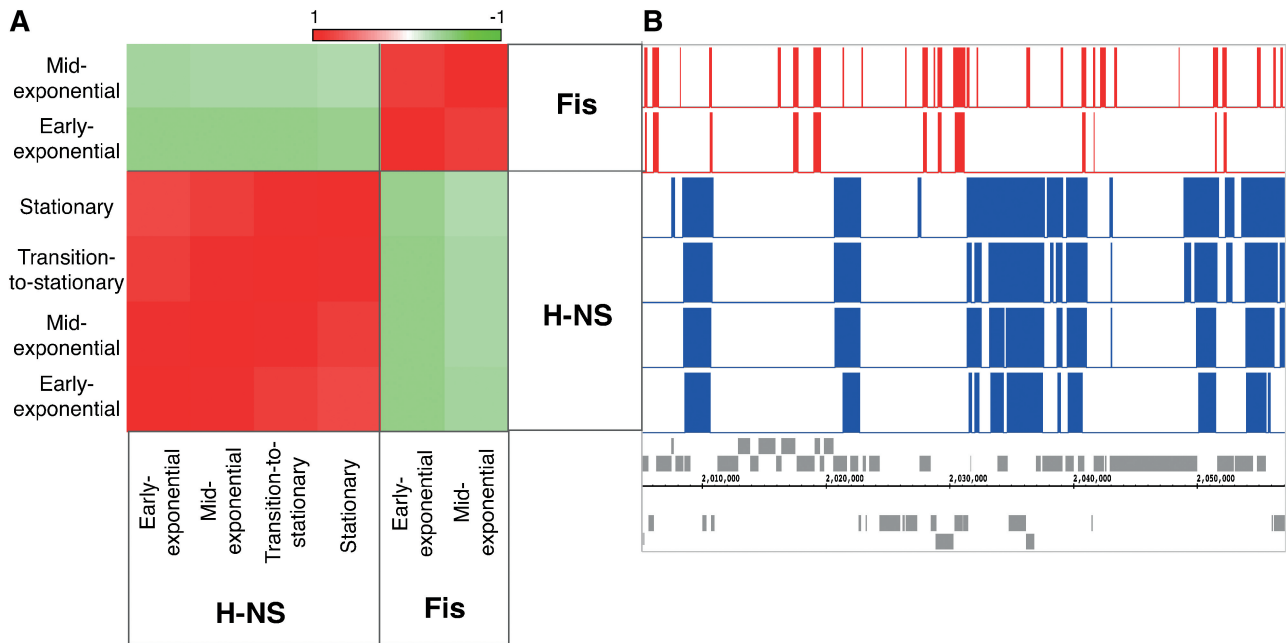
Though the binding profiles are significantly correlated between time-points (Figure 4A), there are specific differences (Supplementary Data 10). For H-NS, the number of

binding regions and genes targeted for binding increase as the cells progress from exponential to stationary phase; this includes both stationary phase-specific binding regions and extension of mid-exponential phase binding regions. For Fis, we observe greater variability in binding between early- and mid-exponential phases than in H-NS ( $\rho = 0.85$  for Fis compared with 0.95 for H-NS, between early- and mid-exponential phases), with more binding in mid-exponential phase. However, we advocate caution in interpreting these results, as they may represent marginal quantitative differences resulting from the thresholds used to call binding events and therefore have limited biological relevance.

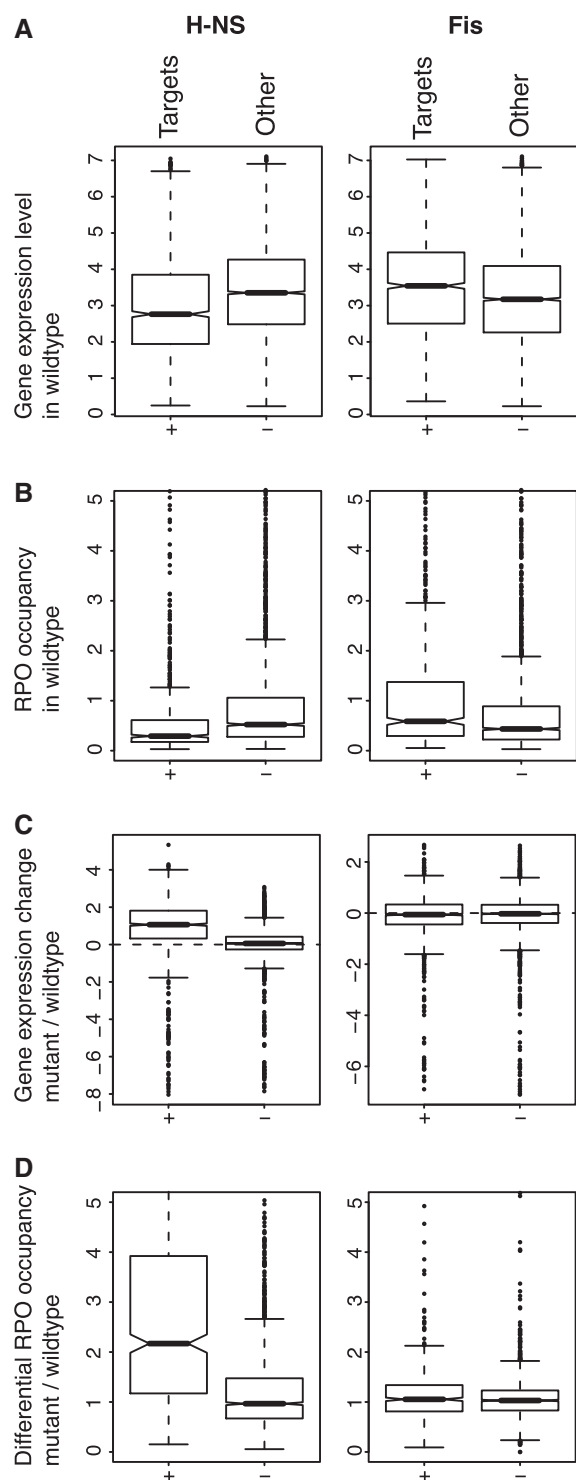
Finally, as mentioned in the section above, there is a negative correlation between H-NS and Fis at each time-point (Figure 4A), suggesting that the binding regions of the two proteins tend to be mutually exclusive.

**Direct, proximal effects of H-NS and Fis binding on gene expression**

*Genes bound by H-NS and Fis show different expression levels in wild-type E. coli.* Having examined the pattern of DNA binding by H-NS and Fis, we investigated whether genes bound by H-NS and Fis showed distinct patterns of gene expression in wild-type *E. coli* cells during mid-exponential phase. Using one-colour experiments on Agilent oligonucleotide microarrays, we found that absolute gene expression levels (which correlate with expression measures derived from RNAseq data) were: (i) lower for genes bound by H-NS than those that are not; and (ii) higher for genes bound by Fis (Figure 5A; Supplementary Data 11). We make consistent



**Figure 4.** Temporal patterns in the binding of H-NS and Fis to the chromosome. (A) Matrix showing the correlation coefficients between the binding profiles of H-NS and Fis across the four time-points studied. Note that Fis is not detectable in transition-to-stationary and stationary phases of growth. Red represents positive and green negative correlation coefficients. (B) Genomic browser (Integrated Genome Browser) view of an  $\sim 50$  kb region of the *E. coli* genome showing H-NS (blue) and Fis (red) binding regions across the four time-points; the grey bars in the bottom panel represent annotated genes on the + (top) and the - (bottom) strands.



**Figure 5.** Protein binding and gene expression and RNA-polymerase occupancy: (A) Gene expression levels in wild-type cells of H-NS (left) and Fis (right) target- (labelled with '+') and non-target ('-') genes. (B) RNA-polymerase occupancy in the body of target and non-target genes. (C) Gene expression fold change (log<sub>2</sub> scale) in the deletion strains (over wild-type) of target and non-target genes. (D) RNA-polymerase occupancy fold change in the deletion strains (over wild-type) of target and non-target genes. All plots here are for the mid-exponential phase.

observations in experiments in which we measure genome-wide RNA-polymerase occupancy during mid-exponential phase using ChIP-seq (Figure 5B). The former observation is in line with the accepted role of H-NS as a global repressor of gene expression (63). The latter, linking Fis binding to higher expression levels, may be consistent with the hypothesis that a branched DNA topology, which is induced by Fis binding, is a chromatin state that is associated with transcriptional activity (15).

We compared our data with a public dataset from Vora and colleagues describing general protein occupancy across the *E. coli* genome (Supplementary Data 12) (43). These authors classified domains of high occupancy into those with high gene expression (hePOD; highly expressed protein occupancy domains) and those that are transcriptionally silent (tsPOD). As expected, we find a strong enrichment for H-NS-bound genes within tsPODs (Fisher's exact test,  $P < 10^{-50}$ ). In contrast to observations by the above authors (made using computational searches of Fis-binding motifs), we find that Fis-bound genes are under-represented within tsPODs (Fisher's exact test,  $P = 9.0 \times 10^{-5}$ ).

Though these results show that there is an association between protein binding and the transcriptional state of the corresponding genes, they do not establish causality. In order to test this *in vivo*, we measured gene expression levels for  $\Delta hns$  and  $\Delta fis$  strains of *E. coli* K12 MG1655, and verified selected results using RT-PCR. In agreement with our observations of expression levels in wild-type strains, more genes are up- than down-regulated in  $\Delta hns$  when compared with the wild-type (971 are up-regulated; 335 are down-regulated in mid-exponential phase; Supplementary Data 13), whereas the contrary is true for  $\Delta fis$  (338 are down-regulated; 160 are up-regulated). In order to investigate whether these effects are proximal on the chromosome to the binding regions of Fis and H-NS, we compared our ChIP-seq-based binding profiles with the genes that are differentially expressed in the mutant strains when compared with the wild-type.

**Global transcriptional repression by H-NS.** A significant proportion of genes that are bound by H-NS display differential up-regulation of gene expression in  $\Delta hns$  during mid-exponential phase: 65% of H-NS-bound genes are differentially expressed compared with only 19% of genes not bound by H-NS (Figure 5C; Supplementary Data 14). Similarly, the RNA-polymerase occupancy in the body of genes bound by H-NS increases significantly in  $\Delta hns$ , again demonstrating increased transcription in the mutant strain (Figure 5D).

Previous genome-scale studies had reached conflicting conclusions on the manner in which H-NS represses transcription. ChIP-chip data for *S. enterica Typhimurium* H-NS by Lucchini and colleagues showed that RNA-polymerase is excluded from H-NS bound regions (8). On the other hand, the work of Grainger *et al.* and Oshima *et al.* showed that the polymerase was bound to 50–65% of H-NS bound sites though the associated genes were transcriptionally inactive (17,18); as a result they proposed that H-NS-mediated repression might generally involve trapping the polymerase at the promoter. We find



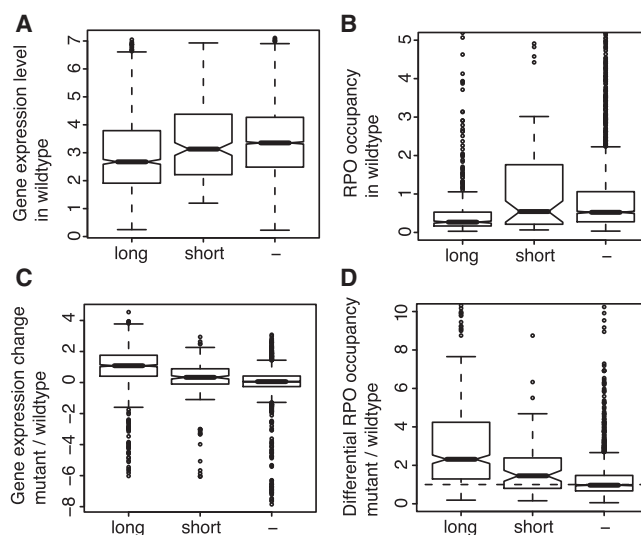
a distinct increase in the RNA-polymerase occupancy *upstream* of operons bound by H-NS in  $\Delta hns$  when compared with the wild-type (Supplementary Data 15), which is concomitant with a corresponding increase in the enzyme's occupancy in the gene body; this suggests that our data support the conclusions of Lucchini *et al.* However, it must be mentioned here that RNA-polymerase trapping by H-NS, though not prevalent in our data, has been experimentally demonstrated at certain promoters (64,65). The differences between the studies, especially with that by Grainger and colleagues, must be interpreted in light of the substantially different numbers of bound genes identified.

In order to extend our analysis further, we performed our DNA microarray-based analysis of gene expression change under all four conditions of growth (Supplementary Data 13 and 14). H-NS has a statistically significant direct impact on gene expression across all phases of growth. However, compared with mid-exponential phase a much smaller proportion of genes bound by H-NS are differentially regulated in  $\Delta hns$  in stationary phase (65% of H-NS bound genes are flagged as differentially expressed in mid-exponential compared with only 26% in stationary phase). This could partly be a consequence of the relatively poor quality of RNA that could be collected from the stationary phase cells, which would lead to the assignment of weaker statistical significance to differential regulation; the total number of genes called as differentially expressed in stationary phase is far less than that in mid-exponential phase (1313 differentially expressed genes in mid-exponential compared with only 400 in stationary phase). Alternatively, there could be a biological basis to this, in which any possible gene expression increase in  $\Delta hns$  is suppressed by other stationary phase-specific factors.

**Differential expression in  $\Delta hns$  is associated with the length of binding regions.** Having described the effect of H-NS binding on gene expression, we now examine the effect of different types of binding ( $L_{H-NS}$ /long and  $S_{H-NS}$ /short) described earlier. Both  $L_{H-NS}$  and  $S_{H-NS}$  operons show a significant tendency to be differentially expressed in  $\Delta hns$  (Figure 6C and D; Supplementary Data 14); however,  $L_{H-NS}$  operons tend to display more differential expression than  $S_{H-NS}$  indicating a greater degree of repression. Further, in the wild-type,  $L_{H-NS}$  genes show lower expression levels than  $S_{H-NS}$  genes (Figure 6A and B, Supplementary Data 11).

To test further whether  $L_{H-NS}$  and  $S_{H-NS}$  genes represent distinct modes of transcriptional repression, as indicated by the above results, we compared their occurrence within tsPODs which represent transcriptionally silent loci (43). We find that  $L_{H-NS}$  genes are enriched within tsPODs, whereas  $S_{H-NS}$  genes are not (Fisher's exact test,  $P = 4.7 \times 10^{-13}$  comparing L and S genes; Supplementary Data 12).

Together, these suggest that global regulation of transcription by H-NS may encompass: (i) transcriptional modulation, typically mild repression, of  $S_{H-NS}$  genes and (ii) 'total' transcriptional 'silencing' of  $L_{H-NS}$  genes, including putative horizontally-acquired genes (20). The



**Figure 6.** Length of binding regions and effect on gene expression. (A) Gene expression levels (log<sub>2</sub> scale) in wild-type; (B) RNA-polymerase occupancy in wild-type; (C) gene expression fold change in  $\Delta hns$ ; (D) RNA-polymerase occupancy fold change in  $\Delta hns$  of genes bound by long and short H-NS binding regions. In (A–D), the distributions for genes not classified as bound by H-NS are provided as a point of reference.

former, given the propensity of the corresponding binding regions to lie within operon-upstream regions, might act like a canonical TF; transcriptional silencing on the other hand involves extensive wrapping of large tracts of the chromosome. Based on the overall distribution of the lengths of H-NS-binding regions, we suggest that the predominant role of H-NS is transcriptional silencing.

**Genes bound by Fis show only limited change in expression in  $\Delta fis$ .** Though the role of H-NS as a transcriptional repressor is well-established, the impact of Fis on gene expression on a genomic scale remains unclear. Given that genes bound by Fis, on average, have higher expression levels in wild-type *E. coli*, one might reasonably expect these genes to be down-regulated in  $\Delta fis$ . Activation of transcription of individual operons, those of stable RNA in particular, by Fis is well-characterized (60,66,67); an inspection of regulatory targets for Fis in the RegulonDB database suggests that it activates more genes than it represses. However, it must be emphasised that activation of gene expression does not fully explain the regulatory roles of Fis as it is a key repressor of several non-essential genes during exponential growth (68–70).

In our study, the large number of genes differentially expressed in  $\Delta fis$  account for only a small proportion of Fis-bound genes (Figure 5C; Supplementary Data 14). We also make a consistent observation in the sequencing-based RNA-polymerase occupancy data for mid-exponential phase (Figure 5D), thus indicating that the above is not an artefact of the array technology. Our results are in agreement with a previously published ChIP-chip study of Fis, which showed differential expression for only about a quarter of bound genes (19). Curiously, despite the general agreement in Fis-binding

regions between early- and mid-exponential phases of growth, there is little overlap between the sets of genes that are differentially expressed in *Δfis* between the two time points; similar observations were made earlier for Fis in *E. coli* (71) and IHF in *Salmonella typhimurium* (72).

These data indicate that deletion of *fis* is not sufficient to cause expression change in most genes that are bound by this protein; this might be because Fis only has a weak role as a TF in these genes, or because these effects are compensated for by other *cis*- and *trans*-acting players which we do not study here.

**Differential expression in *Δfis* is associated with the length, strength and position of Fis binding.** We then investigated whether binding regions associated with the relatively fewer differentially-expressed genes in *Δfis* show any distinctive property. These, when compared with binding regions associated with genes not differentially expressed in *Δfis*, (i) tend to be longer (Figure 7A; Wilcoxon test,  $P = 2.5 \times 10^{-10}$  for mid-exponential phase) and consequently contain more Fis binding motifs (Figure 7B); (ii) have higher A/T content (Figure 7F; Wilcoxon test,  $P < 10^{-50}$ ). Following from the latter (see ‘Variable structures of Fis–DNA complexes’ section), these binding regions also tend to have higher binding signals (Figure 7C and D; Wilcoxon test,  $P = 7.0 \times 10^{-8}$ ), and contain a greater proportion of operon-upstream motifs (Figure 7E; Wilcoxon test,  $P = 3.0 \times 10^{-28}$ ).

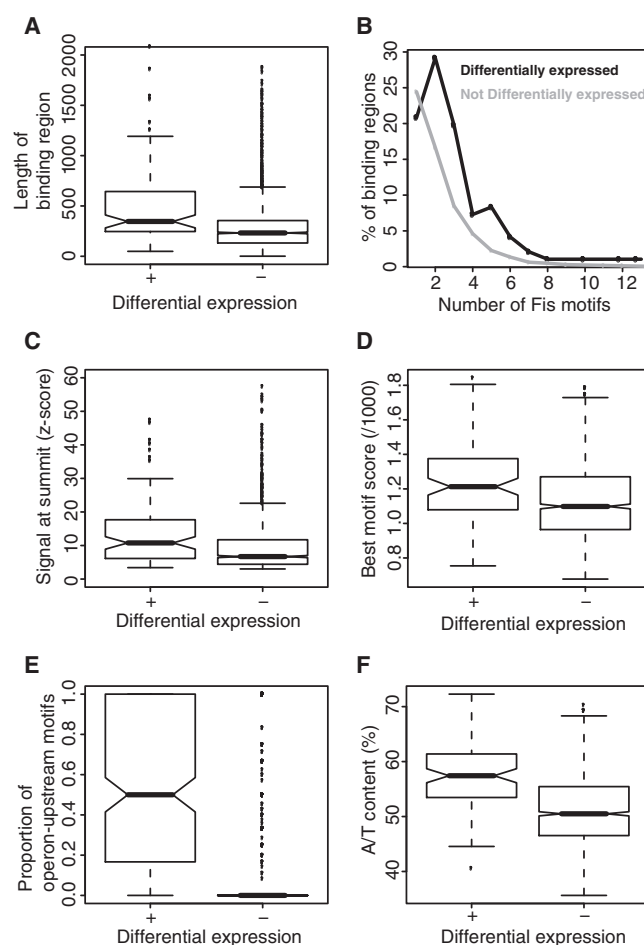
These results indicate that change in expression of a gene bound by Fis might require Fis-binding in multiple tandem copies, possibly nucleated by high-affinity sites at operon-upstream regions. The higher A/T content of binding motifs associated with proximal differential expression suggests that, in accordance with observations made on a molecular scale, DNA-bending by Fis might be required for gene expression control (62). These are exemplified by the *tyrT* promoter which is regulated by three Fis dimers binding and bending the DNA (66,67). However, these features are not predictive of differential expression (Supplementary Data 16), indicating that definitive determinants of gene expression control by Fis are still lacking.

### Indirect and non-proximal effects of H-NS and Fis binding on gene expression

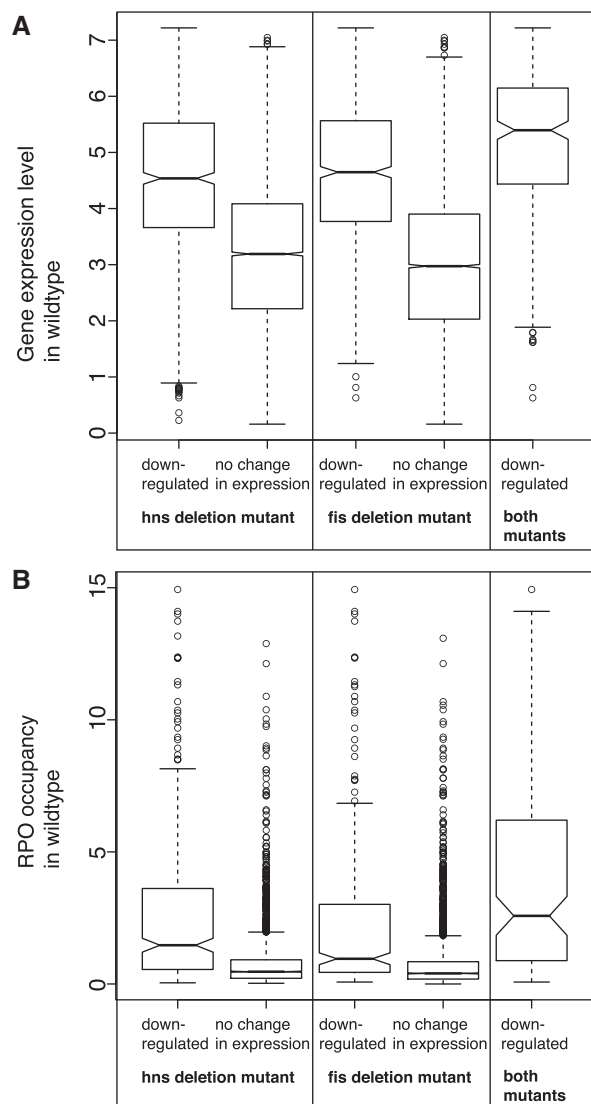
**Down-regulation of highly expressed genes in *Δhns* and *Δfis*.** A large number of genes are down-regulated in both *Δhns* and *Δfis*, a large majority of which are not bound by the NAPs concerned; therefore these effects are likely to be indirect. Genes that are down-regulated in the two deletion strains tend to have higher expression levels than other genes in the wild-type strain (Figure 8). Thus, despite the dissimilarities in the binding of H-NS and Fis, an important minority of their influence on gene expression—especially of highly expressed genes—is shared. This might be a consequence of the impact the two proteins have on the topology of the chromosome—its supercoiled state in particular (73)—which, despite *in vitro* studies on plasmids and phage DNA, is only

beginning to be characterized on a genome-wide scale and at a high resolution (74,75).

Given that genes that are down-regulated in the deletion strains tend to have high expression levels, we sought to mine our data to speculate on how the free RNA-polymerase molecules thus generated are redistributed in the mutants. A significantly higher proportion of genes up-regulated in *Δhns* than in *Δfis* have RNA-polymerase occupancy that are within the top 10% of highly expressed genes (12% of up-regulated genes in *Δhns*, 3% in *Δfis*,  $P = 2.6 \times 10^{-4}$ ). Thus, both deletions lead to fall in expression of highly expressed genes; however, the manner in which the free RNA-polymerase molecules are redistributed may be different between the two. In *Δfis*, these are probably distributed across genes with



**Figure 7.** Fis binding features associated with differential gene expression. (A) Length of binding regions associated with Fis-bound genes which are differentially expressed (+) and not (–) in *Δfis*. (B) Number of Fis binding motifs in binding regions associated with Fis-bound genes which are differentially expressed (black) and not (grey) in *Δfis*. (C) Binding signal at the summit, and (D) scores assigned to the best motif within each binding region associated with Fis-bound genes which are differentially expressed (+) and not (–) in *Δfis*. (E) Distribution of the proportion of operon-upstream motifs in each binding region associated with Fis-bound genes which are differentially expressed (+) and not (–) in *Δfis*. (F) Distribution of the A/T content (50 bp on either side of the summit) of each binding region associated with Fis-bound genes which are differentially expressed (+) and not (–) in *Δfis*.



**Figure 8.** Indirect effects of H-NS and Fis on gene expression. (A) Distributions of expression levels in wild-type during mid-exponential phase of genes which are down-regulated in either *Δhns* or *Δfis* or in both. Also shown are genes which are not differentially expressed in each of the two deletion mutants. (B) As above, but showing RNA-polymerase occupancy instead of gene expression levels.

relatively low expression levels; on the other hand, in *Δhns* this is compensated for by a subset of genes whose repression is relieved by the lack of H-NS (51 of 80 up-regulated operons in the top 10% of genes with the highest RNA-polymerase occupancy in *Δhns* are bound by H-NS).

**Non-proximal effects of H-NS on motility.** An observable phenotype of *Δhns* is loss of motility. These genes are not directly regulated by H-NS, making them targets for studying non-proximal effects of H-NS on gene expression. Though the expression of the transcription factor FlhDC—the master regulator of flagellar gene expression—has been reported to be directly regulated by H-NS (76), we do not find evidence for the same in any of the conditions tested. Instead, we find that 17 of the 26 operons coding for cyclic-di-GMP-metabolising

GGDEF/EAL domain-containing proteins, which regulate the switch between motility and adhesion, are bound by H-NS in at least one of the four conditions; 22 of the 29 such genes are differentially expressed in *Δhns* (Supplementary Data 17). It has already been shown that two GGDEF/EAL proteins that inversely control adhesion through regulating curli biogenesis are regulated by H-NS (59); indeed, we observe binding and regulation of *csgD*—a transcriptional regulator of curli biogenesis—by H-NS under all conditions. Our genome-scale study indicates that H-NS is a global regulator that is positioned at the apex of the c-di-GMP regulatory network controlling motility and adhesion.

*Cascading transcriptional regulatory interactions are responsible for part of non-proximal effects of Fis.* Finally, a large majority of genes bound by Fis show little change in gene expression in *Δfis*. However, the *Δfis* mutation leads to a global change in gene expression during the exponential phases of growth, with over 950 genes differentially expressed in early- or mid-exponential phases of growth. Clearly, most of these gene expression changes are caused by indirect effects. These effects might be mediated by the impact of Fis on the overall chromosome topology. A second, more tractable, effect might be through cascades of transcription factors. To investigate this, we used the transcriptional regulatory network comprising 3254 interactions between 163 TFs and 1450 target genes available in RegulonDB. We find that 37 TFs, including the prolific global regulator CRP, are differentially expressed in *Δfis* in early- or mid-exponential phases of growth. Of the 851 annotated targets of these TFs, 316 (37%) are differentially expressed in *Δfis*; this represents a significant enrichment over other genes of which only ~20% are differentially expressed (Fisher's Exact test,  $P = 5.9 \times 10^{-13}$ ). Of the 37 TFs differentially expressed, only 12 are bound directly by Fis. The regulatory cascade effect described holds even if we were to restrict our analysis to the targets of these 12 TFs (199 of 541 targets are differentially expressed; 37%). Of the remaining 25 TFs, 10 are known direct targets of the Fis-bound TFs. Therefore the expression change of 22 of the 37 TFs can be explained by direct Fis binding or by regulation by Fis-bound TFs.

In summary, a significant proportion of genes are differentially expressed in *Δfis* probably because of the cascading effects of multiple transcription factors.

## DISCUSSION

We have investigated the genome-wide binding of two NAPs, H-NS and Fis, to the *E. coli* K12 MG1655 chromosome using ChIP followed by sequencing of resulting DNA. Though this technique has been extensively adopted in eukaryotic genomics, to our knowledge ours is the first ChIP-Seq experiment for any global bacterial DNA-binding protein. We interpret our data using a combination of deletion strains, microarray-based measurements of gene expression and parallel-sequencing of RNA-polymerase-bound DNA fragments.



### The present study in context: comparison with previous ChIP-chip studies

The binding of both NAPs has been studied on a genome-wide scale using microarrays. Grainger and colleagues studied the binding of H-NS, Fis (and IHF) in mid-exponential phase and expanded the list of genes known to be bound by these proteins (17). In particular, they reported the presence of extensive overlap between the promoters bound by Fis and H-NS, which we do not observe in our conditions. Our observation of a negative correlation between the ChIP-Seq signals for Fis and H-NS is further manifested by the following observations: (i) H-NS binding is enriched in putative horizontally-acquired regions, whereas Fis binding is not; (ii) H-NS targets are enriched in transcriptionally silent protein-occupied DNA domains, whereas Fis-bound genes are under-represented. This difference in observation between our study and that of Grainger and colleagues (17) is caused by the discrepancy between the two in identifying Fis binding regions. Though these differences are surprising, they may be explained in various ways. First, Fis might bind, with a range of affinities, to most of the *E. coli* genome; this may be observed in the higher background in our Fis experiment (Figure 1). Therefore, each study may be sampling a distinct set of bound loci. Second, the experimental conditions are vastly different: our experiments were carried out in rich LB medium without sugar supplements, whereas Grainger *et al.* performed theirs in M9 minimal medium plus fructose. This could have led to substantial differences in Fis binding profiles due to its reported association with catabolite repression and competition with the global transcription factor CRP (53,54). Analysis performed here shows statistically significant overlap between Fis-bound genes and known CRP targets. This suggested association between Fis and CRP targets might be indicative of cooperative or competitive interactions; however given these data, this cannot be substantiated at present. Taken together, there might be substantial differences in Fis function between rich and minimal medium, and in the presence and absence of catabolite repression-inducing sugars. In addition to the above, the following factors might have had relatively minor effects on the results. We used antibodies against the FLAG epitope which had been tagged to the protein of interest, whereas Grainger and colleagues used direct antibodies. The use of the same antibody against three different proteins makes the data from each protein more comparable by eliminating the effect of differential affinities that different antibodies might have towards their target proteins. Though the use of a tag might alter the function of the target protein, microarray analysis of gene expression in the tagged strains show that these effects are insubstantial (Supplementary Data 18). Finally, the lengths of the fragments used for sequencing (~200 bp) and microarray hybridization (500–1000 bp), and therefore the achievable resolution, are generally different (38).

Lucchini and co-workers used a similar low-resolution array to investigate the binding of H-NS to the genome of *S. enterica Typhimurium* (8). The important conclusion of

this study, which was independently demonstrated in the same organism by Navarre and colleagues (58), was the silencing of horizontally-acquired genes by H-NS. They showed that H-NS-binding regions in general exclude RNA-polymerase. Oshima and colleagues identified binding regions of H-NS in *E. coli* using high-resolution microarrays and again showed its effect on horizontally-acquired genes (18). In contrast to the conclusions of Lucchini *et al.* and in agreement with those of Grainger and co-workers (17), these authors identified binding of RNA-polymerase to operon-upstream H-NS-binding regions, though the proximal genes are transcriptionally silent. Our data and analyses do not support this, possibly because of differences between the studies in the sampled binding sites, but are in agreement with those of Lucchini *et al.* Both Lucchini and Navarre have also demonstrated that uncontrolled expression of H-NS-silenced genes can lead to fitness defects (8,58). However, under the conditions used in our study, the wild-type and  $\Delta hns$  have similar growth rates. The difference between our observations might be due to the nature of the genes which are regulated by H-NS in the two organisms. This is reflected in our observation that a majority of H-NS targets *Salmonella* are not conserved in *E. coli* (see section 'Comparison with previously published high-throughput datasets'), in line with the tendency of H-NS to silence horizontally-acquired genes.

The above studies were performed only during mid-exponential phase of growth, though Grainger and co-workers extended theirs to a medium supporting lower growth rates (17). A more recent genome-wide interrogation of H-NS-genome interactions by Noom and colleagues was interpreted, albeit tenuously, in the context of the formation of looped domain boundaries in the *E. coli* and *S. typhimurium* chromosomes (77). These authors performed their study in stationary phase cells, in addition to mid-exponential cells: in agreement with the documented 2-fold decrease in H-NS levels in stationary phase, the authors found that the spacing between adjacent H-NS binding patches doubles in stationary phase. In contrast, we find no evidence for decreased H-NS expression or binding in stationary phase, in agreement with observations made earlier for H-NS in *Salmonella* (9).

Cho and colleagues used high-density genome-tiling microarrays to interrogate the binding of Fis to the *E. coli* genome during mid-exponential growth under aerobic and anaerobic conditions, again in minimal medium (19). They showed that there is little difference in binding profiles between aerobic and anaerobic conditions, a comparison we do not perform. On the other hand, unlike our study they did not investigate multiple time-points during a growth phase. Similar to our conclusions, these authors found little association between Fis binding and differential expression in  $\Delta fis$ . This extends the observations made for another global transcriptional regulator CRP in *E. coli* (41), a large majority of whose binding sites are likely to have little effect on transcription. This led the authors to propose that the primary role of CRP is to structure the chromosome in an as yet uncharacterized manner; its role as a global transcription

factor might be an incidental development. A similar interpretation may be valid for Fis as well.

### Impact of binding characteristics on gene expression

Despite substantial overlap between our study and those of earlier investigations, we extend our interpretation by analysing the association between the nature of binding patches and its influence on gene expression. We show that H-NS binds to significantly longer patches of the chromosome than Fis, in both early- and mid-exponential phases. We speculate that these long binding tracts might include both arms of the plectonemic supercoils and the apical loops that H-NS introduces on the bound DNA (3,10); however, we note that this does not rule out the fact that instead of bridging DNA, H-NS might stiffen the bound DNA at certain sites (78). These long regions of H-NS binding enable transcriptional silencing—displaying greater differential expression in *Δhns* and also showing an enrichment for being present within protein occupancy domains associated with transcriptionally silent loci—whereas shorter patches act as gentler modulators of gene expression. Short H-NS binding regions display a greater preference towards binding to operon-upstream regions than both long H-NS- and Fis-binding regions. This tendency of short H-NS-binding regions to behave more typically like canonical transcription factors than Fis binding regions might explain the relatively greater proximal effect of short H-NS binding patches on gene expression when compared with Fis (40% of genes targeted by short H-NS binding regions are differentially expressed in mid-exponential phase, whereas only 15% of Fis targets are; Fisher's exact test,  $P = 1.6 \times 10^{-8}$ ).

As mentioned above, both our study and that by Cho *et al.* discover that a large majority of strong Fis-binding events are inconsequential from the transcriptional perspective; however, we additionally suggest that the interaction of tandem arrays of Fis molecules to the DNA and possible DNA bending, particularly at operon-upstream regions, might be necessary, though not sufficient, for affecting transcription. Further, we notice that signals in our ChIP-Seq experiments for Fis are weaker than those for H-NS (Figure 1). This observation must be interpreted with caution since the efficiency of immunoprecipitation may depend on the clustering of multiple target proteins on the same chromosomal loci. Additionally, this might also be due to a higher background for Fis, resulting from weak or sporadic binding events across the genome. If this difference is indeed because Fis–DNA interactions in general are weaker and/or more dynamic than H-NS–DNA contacts, it might be responsible for the relatively weaker association between Fis binding and proximal gene expression change.

In contrast to previous studies, we also perform an analysis of the origins of non-proximal effects of the binding of Fis and H-NS to the chromosome. We show a general decrease in the expression of highly expressed transcripts in both the deletion strains, and speculate on the manner in which the RNA-polymerase is redistributed

in these mutants: whereas foci of high transcriptional activity may be lost in *Δfis*, these are replaced by H-NS-bound genes in *Δhns*.

### Perspectives

The main roles of NAPs, particularly in relation to gene expression control, are still under active investigation. Though our study contributes to this field, it leaves several questions, including the following, unanswered. (i) What is the predominant function of Fis-chromosome interactions? (ii) What are the implications, if any, of our observation that, on a genome-wide scale, there is a higher background signal for Fis than H-NS? (iii) What factors definitively link Fis binding to proximal gene expression change?

Finally, we also provide a proof-of-principle study for the use of massively parallel high-throughput sequencing for the analysis of protein–DNA interactions on a genomic scale in bacteria. This is a state-of-the-art technology which affords significantly higher resolution and dynamic range than microarray-based studies. However, there is substantial room for improvement. For example, modifications to the ChIP protocol, which minimize experimental artifacts—including capture of large molecular weight complexes—were proposed very recently (79). Second, from the sequencing perspective, multiplexing techniques are under active development (80). Since 10–15-fold coverage of the genome (compared with ~150-fold obtained in our study) should enable good recovery of binding regions for most bacterial proteins, multiplexing should make ChIP-Seq more economical and therefore prevalent in the field.

### ACCESSION NUMBERS

Array Express E-MTAB-332, Array Express A-MEXP-1866, Array Express E-MEXP-2838, European Nucleotide Archive ERP000280, Array Express E-MTAB-387.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGEMENTS

The authors thank David Grainger, Sacha Lucchini and Nadia Abed for their advice on ChIP protocols. We thank Prof. Stephen Busby and Dr David Grainger for helpful discussion.

### FUNDING

Cambridge Commonwealth Trust; St. John's College, University of Cambridge; Girton College, University of Cambridge (to A.S.N.S.); Spanish Ministry of Science and Innovation (to A.I.P); Biotechnology and Biological Sciences Research Council (BBSRC) grant 'Genomic Analysis of Regulatory Networks for Bacterial Differentiation and Multicellular Behaviour' (to G.M.F.

and N.M.L.); Isaac Newton Trust (to G.M.F); European Molecular Biology Laboratory (EMBL) (to N.M.L.). Funding for open access charge: European Molecular Biology Laboratory.

*Conflict of interest statement.* None declared.

## REFERENCES

- Browning, D.F. and Busby, S.J. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.*, **2**, 57–65.
- Martínez-Antonio, A. and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.*, **6**, 482–489.
- Dillon, S.C. and Dorman, C.J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.*, **8**, 185–195.
- Dorman, C.J. and Kane, K.A. (2009) DNA bridging and antibridging: a role for bacterial nucleoid-associated proteins in regulating the expression of laterally acquired genes. *FEMS Microbiol. Rev.*, **33**, 587–592.
- Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.*, **181**, 6361–6370.
- Azam, T.A. and Ishihama, A. (1999) Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J. Biol. Chem.*, **274**, 33105–33113.
- Dorman, C.J. (2007) H-NS, the genome sentinel. *Nat. Rev. Microbiol.*, **5**, 157–161.
- Lucchini, S., Rowley, G., Goldberg, M.D., Hurd, D., Harrison, M. and Hinton, J.C. (2006) H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathogens*, **2**, e81.
- Hinton, J.C., Santos, D.S., Seirafi, A., Hulton, C.S., Pavitt, G.D. and Higgins, C.F. (1992) Expression and mutational analysis of the nucleoid-associated protein H-NS of *Salmonella typhimurium*. *Mol. Microbiol.*, **6**, 2327–2337.
- Dame, R.T., Noom, M.C. and Wuite, G.J. (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature*, **444**, 387–390.
- Maurer, S., Fritz, J. and Muskhelishvili, G. (2009) A systematic in vitro study of nucleoprotein complexes formed by bacterial nucleoid-associated proteins revealing novel types of DNA organization. *J. Mol. Biol.*, **387**, 1261–1276.
- Finkel, S.E. and Johnson, R.C. (1992) The Fis protein: it's not just for DNA inversion anymore. *Mol. Microbiol.*, **6**, 3257–3265.
- Schneider, R., Travers, A. and Muskhelishvili, G. (1997) FIS modulates growth phase-dependent topological transitions of DNA in *Escherichia coli*. *Mol. Microbiol.*, **26**, 519–530.
- Gama-Castro, S., Jiménez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Peñaloza-Spinola, M.I., Contreras-Moreira, B., Segura-Salazar, J., Muñoz-Rascado, L., Martínez-Flores, I., Salgado, H. et al. (2008) RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.*, **36**, D120–D124.
- Schneider, R., Lurz, R., Lüder, G., Tolksdorf, C., Travers, A. and Muskhelishvili, G. (2001) An architectural role of the *Escherichia coli* chromatin protein FIS in organising DNA. *Nucleic Acids Res.*, **29**, 5107–5114.
- Janga, S.C., Salgado, H. and Martínez-Antonio, A. (2009) Transcriptional regulation shapes the organization of genes on bacterial chromosomes. *Nucleic Acids Res.*, **37**, 3680–3688.
- Grainger, D.C., Hurd, D., Goldberg, M.D. and Busby, S.J. (2006) Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.*, **34**, 4642–4652.
- Oshima, T., Ishikawa, S., Kurokawa, K., Aiba, H. and Ogasawara, N. (2006) *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res.: An Int. J. Rapid Publ. Reports Genes Genomes*, **13**, 141–153.
- Cho, B., Knight, E.M., Barrett, C.L. and Pálsson, B.Ø. (2008) Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. *Genome Res.*, **18**, 900–910.
- Stoebel, D.M., Free, A. and Dorman, C.J. (2008) Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology*, **154**, 2533–2545.
- Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A. et al. (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods*, **4**, 651–657.
- Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.
- Liu, E.T., Pott, S. and Huss, M. (2010) Q&A: ChIP-seq technologies and the study of gene regulation. *BMC Biol.*, **8**, 56.
- König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.*, **17**, 909–915.
- Preti, M., Ribeyre, C., Pascali, C., Bosio, M.C., Cortelazzi, B., Rougemont, J., Guarnera, E., Naef, F., Shore, D. and Dieci, G. (2010) The telomere-binding protein Tbf1 demarcates snoRNA gene promoters in *Saccharomyces cerevisiae*. *Mol. Cell*, **38**, 614–620.
- Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martínez-Jiménez, C.P., Mackay, S. et al. (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, **328**, 1036–1040.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
- Güell, M., van Noort, V., Yus, E., Chen, W., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kühner, S. et al. (2009) Transcriptome complexity in a genome-reduced bacterium. *Science*, **326**, 1268–1271.
- Harris, S.R., Feil, E.J., Holden, M.T., Quail, M.A., Nickerson, E.K., Chantratita, N., Gardete, S., Tavares, A., Day, N., Lindsay, J.A. et al. (2010) Evolution of MRSA during hospital transmission and intercontinental spread. *Science*, **327**, 469–474.
- Passalacqua, K.D., Varadarajan, A., Ondov, B.D., Okou, D.T., Zwick, M.E. and Bergman, N.H. (2009) Structure and complexity of a bacterial transcriptome. *J. Bacteriol.*, **191**, 3203–3211.
- Perkins, T.T., Kingsley, R.A., Fookes, M.C., Gardner, P.P., James, K.D., Yu, L., Assefa, S.A., He, M., Croucher, N.J., Pickard, D.J. et al. (2009) A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genet.*, **5**, e1000569.
- Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R. et al. (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, **464**, 250–255.
- Vivancos, A.P., Güell, M., Dohm, J.C., Serrano, L. and Himmelbauer, H. (2010) Strand-specific deep sequencing of the transcriptome. *Genome Res.*, **20**, 989–999.
- Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C. and Vogel, J. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.*, **4**, e1000163.
- Lun, D.S., Sherid, A., Weiner, B., Sherman, D.R. and Galagan, J.E. (2009) A blind deconvolution approach to high-resolution mapping of transcription factor binding sites from ChIP-seq data. *Genome Biol.*, **10**, R142.
- Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA*, **97**, 6640–6645.
- Uzzau, S., Figueroa-Bossi, N., Rubino, S. and Bossi, L. (2001) Epitope tagging of chromosomal genes in *Salmonella*. *Proc. Natl Acad. Sci. USA*, **98**, 15264–15269.



38. Grainger, D.C., Overton, T.W., Reppas, N., Wade, J.T., Tamai, E., Hobman, J.L., Constantinidou, C., Struhl, K., Church, G. and Busby, S.J. (2004) Genomic studies with *Escherichia coli* MelR protein: applications of chromatin immunoprecipitation and microarrays. *J. Bacteriol.*, **186**, 6938–6943.
39. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D3560.
40. Keseler, I.M., Bonavides-Martínez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R.P., Johnson, D.A., Krummenacker, M., Nolan, L.M., Paley, S., Paulsen, I.T. *et al.* (2009) EcoCyc: a comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res.*, **37**, D464–D470.
41. Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J. and Busby, S.J. (2005) Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc. Natl Acad. Sci. USA*, **102**, 17693–17698.
42. Vernikos, G.S. and Parkhill, J. (2006) Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics*, **22**, 2196–2203.
43. Vora, T., Hottes, A.K. and Tavazoie, S. (2009) Protein occupancy landscape of a bacterial genome. *Mol. Cell*, **35**, 247–253.
44. Moreno-Hagelsieb, G. and Latimer, K. (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*, **24**, 319–324.
45. David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W. and Steinmetz, L.M. (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA*, **103**, 5320–5325.
46. Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M. and Gerstein, M.B. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat. Biotechnol.*, **27**, 66–75.
47. Barski, A. and Zhao, K. (2009) Genomic location analysis by ChIP-Seq. *J. Cell. Biochem.*, **107**, 11–18.
48. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
49. Bozdech, Z., Zhu, J., Joachimiak, M.P., Cohen, F.E., Pulliam, B. and DeRisi, J.L. (2003) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol.*, **4**, R9.
50. Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A. and Smyth, G.K. (2007) A comparison of background correction methods for two-colour microarrays. *Bioinformatics*, **23**, 2700–2707.
51. Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A. and Vingron, M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18**(Suppl. 1), S96–S104.
52. Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article3.
53. Browning, D.F., Grainger, D.C., Beatty, C.M., Wolfe, A.J., Cole, J.A. and Busby, S.J. (2005) Integration of three signals at the *Escherichia coli* *nrf* promoter: a role for Fis protein in catabolite repression. *Mol. Microbiol.*, **57**, 496–510.
54. Galán, B., Manso, I., Kolb, A., García, J.L. and Prieto, M.A. (2008) The role of FIS protein in the physiological control of the expression of the *Escherichia coli* meta-hpa operon. *Microbiology*, **154**, 2151–2160.
55. Dillon, S.C., Cameron, A.D., Hokamp, K., Lucchini, S., Hinton, J.C. and Dorman, C.J. (2010) Genome-wide analysis of the H-NS and Sfh regulatory networks in *Salmonella* Typhimurium identifies a plasmid-encoded transcription silencing mechanism. *Mol. Microbiol.*, **76**, 1250–1265.
56. Bouffartigues, E., Buckle, M., Badaut, C., Travers, A. and Rimsky, S. (2007) H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat. Struct. Mol. Biol.*, **14**, 441–448.
57. Lang, B., Blot, N., Bouffartigues, E., Buckle, M., Geertz, M., Gualerzi, C.O., Mavathur, R., Muskhelishvili, G., Pon, C.L., Rimsky, S. *et al.* (2007) High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Res.*, **35**, 6330–6337.
58. Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S.J. and Fang, F.C. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science*, **313**, 236–238.
59. Weber, H., Pesavento, C., Possling, A., Tischendorf, G. and Hengge, R. (2006) Cyclic-di-GMP-mediated signalling within the sigma network of *Escherichia coli*. *Mol. Microbiol.*, **62**, 1014–1034.
60. Travers, A., Schneider, R. and Muskhelishvili, G. (2001) DNA supercoiling and transcription in *Escherichia coli*: The FIS connection. *Biochimie*, **83**, 213–217.
61. Mitchison, G. (2005) The regional rule for bacterial base composition. *Trends Genet.*, **21**, 440–443.
62. Pan, C.Q., Finkel, S.E., Cramton, S.E., Feng, J.A., Sigman, D.S. and Johnson, R.C. (1996) Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J. Mol. Biol.*, **264**, 675–695.
63. Fang, F.C. and Rimsky, S. (2008) New insights into transcriptional regulation by H-NS. *Curr. Opin. Microbiol.*, **11**, 113–120.
64. Dame, R.T., Wyman, C., Wurm, R., Wagner, R. and Goosen, N. (2002) Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the *rrnB* P1. *J. Biol. Chem.*, **277**, 2146–2150.
65. Schröder, O. and Wagner, R. (2000) The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. *J. Mol. Biol.*, **298**, 737–748.
66. Pemberton, I.K., Muskhelishvili, G., Travers, A.A. and Buckle, M. (2002) FIS modulates the kinetics of successive interactions of RNA polymerase with the core and upstream regions of the *tyrT* promoter. *J. Mol. Biol.*, **318**, 651–663.
67. Muskhelishvili, G., Travers, A.A., Heumann, H. and Kahmann, R. (1995) FIS and RNA polymerase holoenzyme form a specific nucleoprotein complex at a stable RNA promoter. *EMBO J.*, **14**, 1446–1452.
68. Browning, D.F., Cole, J.A. and Busby, S.J. (2008) Regulation by nucleoid-associated proteins at the *Escherichia coli* *nir* operon promoter. *J. Bacteriol.*, **190**, 7258–7267.
69. Grainger, D.C., Goldberg, M.D., Lee, D.J. and Busby, S.J. (2008) Selective repression by Fis and H-NS at the *Escherichia coli* *dps* promoter. *Mol. Microbiol.*, **68**, 1366–1377.
70. Squire, D.J., Xu, M., Cole, J.A., Busby, S.J. and Browning, D.F. (2009) Competition between NarL-dependent activation and Fis-dependent repression controls expression from the *Escherichia coli* *yeaR* and *ogt* promoters. *Biochem. J.*, **420**, 249–257.
71. Bradley, M.D., Beach, M.B., de Koning, A.P., Pratt, T.S. and Osuna, R. (2007) Effects of Fis on *Escherichia coli* gene expression during different growth stages. *Microbiology*, **153**, 2922–2940.
72. Mangan, M.W., Lucchini, S., Danino, V., Cróinin, T.O., Hinton, J.C. and Dorman, C.J. (2006) The integration host factor (IHF) integrates stationary-phase and virulence gene expression in *Salmonella enterica* serovar Typhimurium. *Mol. Microbiol.*, **59**, 1831–1847.
73. Hardy, C.D. and Cozzarelli, N.R. (2005) A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Mol. Microbiol.*, **57**, 1636–1652.
74. Blot, N., Mavathur, R., Geertz, M., Travers, A. and Muskhelishvili, G. (2006) Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep.*, **7**, 710–715.
75. Peter, B.J., Arsuaga, J., Breier, A.M., Khodursky, A.B., Brown, P.O. and Cozzarelli, N.R. (2004) Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol.*, **5**, R87.
76. Soutourina, O., Kolb, A., Krin, E., Laurent-Winter, C., Rimsky, S., Danchin, A. and Bertin, P. (1999) Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the

- cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon. *J. Bacteriol.*, **181**, 7500–7508.
77. Noom, M.C., Navarre, W.W., Oshima, T., Wuite, G.J. and Dame, R.T. (2007) H-NS promotes looped domain formation in the bacterial chromosome. *Curr. Biol.*, **17**, R913–R914.
78. Liu, Y., Chen, H., Kenney, L.J. and Yan, J. (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev.*, **24**, 339–344.
79. Waldminghaus, T. and Skarstad, K. (2010) ChIP on Chip: surprising results are often artifacts. *BMC Genomics*, **11**, 414.
80. Lefrançois, P., Euskirchen, G.M., Auerbach, R.K., Rozowsky, J., Gibson, T., Yellman, C.M., Gerstein, M. and Snyder, M. (2009) Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing. *BMC Genomics*, **10**, 37.